REPORT DOCUMENTATION PAGE

3

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
	Technical Report	-

4. TITLE AND SUBTITLE
Study of time reversal in complex systems

5a. CONTRACT NUMBER
W911NF-13-1-0390

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611103

6. AUTHORS
Brenden C. Roberts

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
University of California - Davis
1850 Research Park Drive
Suite 300
Davis, CA 95618 -6153

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
63822-EG-MUR.8

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
In this paper we present conclusions of a study of irreversibility in physical processes, using the conceptual formalism of the ?-machine. The causal irreversibility is examined, in particular for the class of Flower processes, and found to be attributable to contributions from the transi- tion probability distributions as well as the causal state topology of a process. The topological irreversibility is a prerequisite for causal irreversibility, and is examined in more detail in the framework of topological ?-machines. Detailed study of the mechanisms involved in time-reversal of the particular model class of semi-periodic deterministic finite automata allows characteriza- tion of the

15. SUBJECT TERMS
computational mechanics, epsilon-machine, casual irreversibility

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			James Crutchfield
UU	UU	UU	UU		19b. TELEPHONE NUMBER 530-752-0060

Report Title

Study of time reversal in complex systems

ABSTRACT

In this paper we present conclusions of a study of irreversibility in physical processes, using the conceptual formalism of the ?-machine. The causal irreversibility is examined, in particular for the class of Flower processes, and found to be attributable to contributions from the transi- tion probability distributions as well as the causal state topology of a process. The topological irreversibility is a prerequisite for causal irreversibility, and is examined in more detail in the framework of topological ?-machines. Detailed study of the mechanisms involved in time-reversal of the particular model class of semi-periodic deterministic finite automata allows characteriza- tion of the topological source of irreversibility in this context.

Study of time reversal in complex systems

Brenden C. Roberts*
Complexity Sciences Center, Department of Physics, UC Davis, Davis, CA 95616
Advisors: Dr. Jim Crutchfield, Dr. John Mahoney

January 2, 2014

Abstract

In this paper we present conclusions of a study of irreversibility in physical processes, using the conceptual formalism of the ϵ-machine. The causal irreversibility is examined, in particular for the class of Flower processes, and found to be attributable to contributions from the transition probability distributions as well as the causal state topology of a process. The topological irreversibility is a prerequisite for causal irreversibility, and is examined in more detail in the framework of topological ϵ-machines. Detailed study of the mechanisms involved in time-reversal of the particular model class of semi-periodic deterministic finite automata allows characterization of the topological source of irreversibility in this context. The work was performed as part of the 2013 UC Davis physics REU program.

Contents

*Department of Physics & Astronomy, Clemson University, Clemson, SC 29634

1 Introduction to complexity sciences

Arguably the primary goal of science is a simple one: to describe the behavior of physical systems. The breadth of this endeavor is immense, and necessitates the development of conceptual models which replicate only those features of a system of interest for a particular purpose. Accordingly, one's goals are crucial to devising or choosing a model, and the utility of any model is determined by the granularity with which one studies a system. The field of complexity sciences is reflective of this. Studying complex systems requires determining the appropriate level of abstraction, at which one can form models comprising distinct elements that collectively exhibit complex emergent behavior. This framework is analogous to the process in thermodynamics by which individual particles behaving according to microscopic force laws are described by concepts like temperature and pressure, which can only be interpreted on a macroscopic scale. Indeed, the notions behind some thermodynamic concepts, like entropy, can be applied in a much broader context. Accordingly, complexity theory seeks to describe the behavior of general composite systems via modeling of the constituents.

Here we place particular emphasis on the case of time evolution of dynamical systems. We define a system by specifying its components and their interactions, identify a quantity that is representative of the state of the system as a whole, and "measure" it as the system evolves over time, although in many cases the quantity is an invariant. The reverse can also be done, where the measurement results are available and the structure of the system is in question. In this case, the ideal result is one that represents the data arising from a system with as few prior assumptions as is possible.

1.1 Presentations of processes

As we are concerned with systems which are evolving in time, and wish to make a comment on the nature of this evolution, the concept of the passage of time is important in motivating a particular construction. The following definitions and notation are taken from references [2], [3], and [4]. Consider some system (henceforth, process $\mathcal{P}$) and an instrument which, with some sampling rate, makes discrete observations of the process, yielding a possibly bi-infinite stream of data:

$$\ldots X_{-3}X_{-2}X_{-1}X_0X_1X_2X_3\ldots$$

Each of the X_i is a random variable representing the data point recorded at time $t = t_i$. Assume the X_i to be discrete random variables which are in general dependent, all of which have probability mass function f_{X_i}. We construct $\mathcal{X} = \bigcup_i \mathrm{im}(f_{X_i})$. That is, $\mathcal{X}$ is the set of all possible measurement outcomes, or the *alphabet*. We refer to any particular realization of measurement data as a sequence of $x_i \in \mathcal{X}$.

A good presentation of this data is one which captures the "macroscopic" features of the process in a more compact manner than the stream of data itself. We also desire predictive capacity in the model. To that end, we interpret the process as a channel transmitting information about the system's history to its future states. We partition the sequence by defining the past $\overleftarrow{X} = \ldots X_{-3}X_{-2}X_{-1}$ and future $\overrightarrow{X} = X_0X_1X_2X_3$. Any particular past $\overleftarrow{x} = \ldots x_{-3}x_{-2}x_{-1}$ is referred to as a *history*. Predictive capacity can now be more precisely defined as the ability to define a probability distribution $\Pr(\overrightarrow{X}|\overleftarrow{x})$ over the space of all possible futures $\overrightarrow{X}$, conditioned on a given history $\overleftarrow{x}$.

1.2 ϵ-machines

To construct such a model of a structured process, we use the histories themselves. We use the equivalence relation that, given our predictive model, $\overleftarrow{x} \sim \overleftarrow{x}'$ if $\Pr(\overrightarrow{X}|\overleftarrow{x}) = \Pr(\overrightarrow{X}|\overleftarrow{x}')$. That is, the equivalence classes are composed of those histories that result in identical predictions. These equivalence classes are referred to as *causal states*, the set of which, $\mathcal{S}$, is a partitioning of the space of histories. Causal states are optimally predictive; that is, by definition, knowledge of the causal state of a process is equivalent to knowing the entire history of the process.

Any process with a history $\overleftarrow{x} = \ldots x_{-3}x_{-2}x_{-1}$ can thus be said to be in a causal state S_0. The next symbol observed, x_0, when added to the history may cause a change in the causal state of the system, taking S_0 to S_1. In this case we refer to a *transition* in the system between causal states upon observation of the symbol. We can construct a set of transition matrices $\mathcal{T} = \{T^x : x \in \mathcal{X}\}$ comprised of a $|\mathcal{S}| \times |\mathcal{S}|$ transition matrix T^x corresponding to each symbol that could be observed. The element T^x_{ij} of a particular matrix specifies the probability of a transition from causal state i to state j upon observing symbol x. Inherent in this description is an assumption that the set of causal states is finite, or requires application of Zorn's lemma. Here we deal only with finite $\mathcal{S}$. The set of causal states with the transition matrices of the process $\mathcal{P}$ comprise the *ϵ-machine* $\mathcal{M}(\mathcal{P})$.

We assume a process to be stationary, giving us freedom in time indexing. We will define X_0 to coincide with the "present" (formally, the first symbol of the future), and, notationally, we let the state S_0 transition to S_1 upon observation of x_1.

Because they carry the same information as a full process history, causal states are Markovian [4]. However, an ϵ-machine is not a Markov chain because the states of the system themselves are not observable; only the symbols are. This property corresponds to a *hidden Markov model*, or HMM. It suggests a graphical representation, which is presented in figure 1 for dummy processes illustrating the difference between the two Markov models.

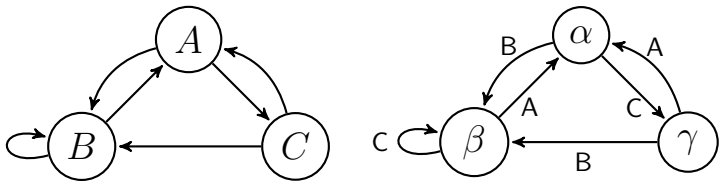


Figure 1: Examples of a state-output Markov chain (a) and hidden Markov model (b) which have the same alphabet $\mathcal{X} = \{A, B, C\}$

The symbol-independent transition matrix can be obtained by integrating $T = \sum_{x\in\mathcal{X}} f(x)T^x$. We may apply T to a $|\mathcal{S}|$-length row vector containing the probability of finding the system in each state. Let π be a left eigenvector of T; then $\pi T = \lambda\pi$ for some λ. But π and $\lambda\pi$ exist in probability space, constraining that their entries must sum to 1. Thus the eigenvalue $\lambda = 1$ and the eigenvector equation is simply $\pi T = \pi$. The π vector solution to this equation is the *stationary state distribution* of the process, the asymptotic limit of the likelihood of finding the system in each state.

Finally, we note that the class of ϵ-machine s is a subset of that of HMMs, as an ϵ-machine must possess two additional properties [3]. First is *unifilarity*, or the property that each symbol uniquely determines a causal state, given the current state. Graphically, this corresponds to no more than one arrow per output symbol originating from each state. The second property is that an ϵ-machine is minimal; there is no simpler HMM that gives identical predictions. To recapitulate,

the ϵ-machine $\mathcal{M}(\mathcal{P})$ is the unique (proof given in [6]) minimal unifilar HMM of a process $\mathcal{P}$, which drastically simplifies the difficulty of calculating invariants of the process and provides optimal predictive capacity.

1.3 Characterizing information-theoretic properties

The ϵ-machine formalism makes explicit the derivation of several invariant properties we wish to define for a given process. These rely on the concept of Shannon entropy. The Shannon entropy is a property of a random variable X that measures the expected value of the "surprise" of its outcomes. The degree of surprise, or information, of some particular outcome i with probability p_i is defined as $I_i = -\log_2 p_i$ [1]. The expectation value, denoted H, is given by

$$H[X] = \int_x f_X(x) I(x) = \sum_{x \in \mathcal{X}} p_x(-\log_2 p_x) \tag{1}$$

We wish to extend this notion to the more general case of a process, rather than a single random variable, via its ϵ-machine. Henceforth we will use lg to refer to the base-2 logarithm.

The first quantity obtained is the statistical complexity C_μ, which we interpret as a measure of the degree of structure of a process, or the amount of historical information communicated by the causal states. It is accordingly measured in bits and is a straightforward generalization of the Shannon entropy. Instead of the entropy of the observed symbols, it is the entropy of the causal states of a process:

$$C_\mu = H[\mathcal{S}] = -\sum_{S \in \mathcal{S}} \Pr(S) \lg \Pr(S) \tag{2}$$

The probability distribution over the causal states is taken from the stationary state vector π, as described in section 1.2.

How do we define the entropy contained in the symbols themselves? This is accomplished through the *entropy rate* h_μ, which is given in units of bits per symbol and is likewise an invariant of a process. It measures the rate at which new information is generated by the symbols, or its degree of stochasticity. This is somewhat less straightforward to derive, as it requires an asymptotic limit to eliminate the vagaries of any finite sequence of symbols. The ϵ-machine formalism simplifies this process. Because infinite-length histories are reduced to a set of causal states $\mathcal{S}$, we calculate the entropy of the symbols conditioned on the set of states; the unifilarity property discussed in section 1.2 ensures that this definition coincides with h_μ:

$$\begin{aligned} h_\mu = H[\mathcal{X}|\mathcal{S}] &= -\sum_{S \in \mathcal{S}} \Pr(S) \sum_{\substack{x \in \mathcal{X} \\ S' \in \mathcal{S}}} \Pr(x|S') \ \lg \Pr(x|S') \\ &= -\sum_{S} \Pr(S) \sum_{x, S'} T^{(x)}_{SS'} \lg T^{(x)}_{SS'} \end{aligned} \tag{3}$$

Examples of these properties, as well as introductions to some of the canonical ϵ-machines, are given in the highly-recommended Appendix A. Also note that, while it is not discussed here, the concept of excess entropy **E** is a similar quantity (the mutual information between the past and future) that is foundational to this formalism [2].

2 Quantifying irreversible behavior

We mentioned earlier that a stationary process allows one freedom with regard to shifting the origin of time-indexing. We now investigate taking another liberty by scanning the time series in the time-reversed direction. We may rederive the fundamentals of our analysis, instead using the future to retrodict the past. This gives a new set of retrodictive causal states $\mathcal{S}^-$, along with transition matrices $\mathcal{T}^-$ comprising the retrodictive ϵ-machine $\mathcal{M}^-$. Henceforth we will also refer to the forward-time versions of these with a superscript: $\mathcal{M}^+$.

It is a result of [3] that the entropy rate h_μ of a time-reversed process is the same as that of the forward process, that $H[X_0|S_{-1}] = H[X_{-1}|S_0]$. This corresponds to the intuitive notion that the amount of information that transmitted by a process is independent of the direction of time. The same is not true of the statistical complexity: many simple counterexamples exist showing that $C_\mu^+ \neq C_\mu^-$. Here we present the Random Noisy Copy (RnC) Process as an example in order to demonstrate the process of time reversal of an ϵ-machine.

2.1 Time reversal

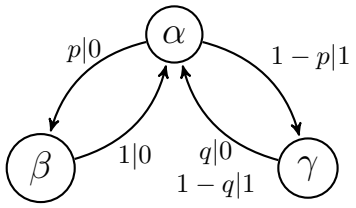


Figure 2: The random noisy copy (RnC) ϵ-machine

Reversal of an ϵ-machine $\mathcal{M}^+$ is a central process to the topic of this report, but one that can be accomplished algorithmically. We will perform the operation once here on the RnC process in order to convey the intuition. It is performed in two steps, denoted by two maps: $\mathcal{T}$, the *time-reversal operator* and $\mathcal{U}$, which computes the *mixed-state presentation* in order to restore unifilarity. The RnC process, in brief, generates symbols in couplets with an alphabet $\mathcal{X} = \mathbb{Z}/2$. The first symbol in each couplet is a 0 with probability p. If this is the case, the next symbol is also a 0 with certainty. If the first symbol is a 1, the next has a probability q of switching to a 0; otherwise it is also a 1. Thus, there are three causal states: the "reset" position, after a couplet is completed, and one state each for the initial symbol being a 0 or 1. These states are $\mathcal{S} = \{\alpha, \beta, \gamma\}$, respectively. Figure 2 gives the visual ϵ-machine. From this we see that

$$T = \begin{bmatrix} 0 & p & 1-p \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \qquad \pi = \tfrac{1}{2}\begin{bmatrix} 1 & p & 1-p \end{bmatrix}$$

$$T^{(0)} = \begin{bmatrix} 0 & p & 0 \\ 1 & 0 & 0 \\ q & 0 & 0 \end{bmatrix} \quad T^{(1)} = \begin{bmatrix} 0 & 0 & 1-p \\ 0 & 0 & 0 \\ 1-q & 0 & 0 \end{bmatrix}$$

Time-reversing the ϵ-machine is a straightforward operation on the transition matrices but is not guaranteed to generate another ϵ-machine. We denote the states of the time-reversed HMM as $\mathcal{R}$

rather than $\mathcal{S}$:

$$\begin{aligned}\widetilde{T}^{(x)}_{R'R} &= \Pr(X = x, R|R') \\ &= T^{(x)}_{RR'} \frac{\Pr(R)}{\Pr(R')}\end{aligned} \tag{4}$$

That is, the time-reversed transition matrices are simply the transposes of those of the forward-time presentation, normalized by the ratios of the stationary state distributions, which are equivalent to the ratios of the stationary causal state distributions ($\widetilde{\pi} = \pi$). This specifies $\mathcal{T}(\mathcal{M}^+) = \widetilde{\mathcal{M}}^+$, which has the same number of causal states as $\mathcal{M}^+$. Applying $\mathcal{T}$ to the RnC gives

$$\widetilde{T}^{(0)} = \begin{bmatrix} 0 & p & q(1-p) \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \widetilde{T}^{(1)} = \begin{bmatrix} 0 & 0 & (1-q)(1-p) \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

We are now ready to apply the mixed-state presentation $\mathcal{U}(\widetilde{\mathcal{M}}^+)$ to obtain an ϵ-machine giving the time-reversed model $\mathcal{M}^-$. The first step is to begin with the stationary state π, a possibly transient causal state that can be viewed as representing a lack of knowledge about the state. This is the state of the system before any symbols ore observed. In general, we will denote the system state after observing word w as $\nu(w)$. After making an observation, we are now in one of two states: $\nu(0)$ or $\nu(1)$. These are simply created by evolving the stationary state by the appropriate transition matrix and normalizing:

$$\nu(0) = \tfrac{\pi\widetilde{T}^{(0)}}{|\pi\widetilde{T}^{(0)}|} \quad \nu(1) = \tfrac{\pi\widetilde{T}^{(1)}}{|\pi\widetilde{T}^{(1)}|} \tag{5}$$

Where $|\cdot|$ denotes the 1-norm of the stochastic vector. Each of these represents a causal state of the ϵ-machine, which may or may not be transient. The quantities $|\pi\widetilde{T}^{(0)}|$ and $|\pi\widetilde{T}^{(1)}|$ represent the possibility of transition to each state from the initial stationary state. The next step in the process is to prepend the next symbols: $\nu(0)$ leads to $\nu(10)$ and $\nu(00)$, and $\nu(1)$ to $\nu(01)$ and $\nu(11)$. To calculate the resultant vectors for each of these, the formula given in equation 5 can be iterated:

$$\nu(00) = \tfrac{\nu(0)\widetilde{T}^{(0)}}{|\nu(0)\widetilde{T}^{(0)}|} \quad \nu(10) = \tfrac{\nu(0)\widetilde{T}^{(1)}}{|\nu(0)\widetilde{T}^{(1)}|}$$

$$\nu(01) = \tfrac{\nu(1)\widetilde{T}^{(0)}}{|\nu(1)\widetilde{T}^{(0)}|} \quad \nu(11) = \tfrac{\nu(1)\widetilde{T}^{(1)}}{|\nu(1)\widetilde{T}^{(1)}|}$$

Again, the denominators represent the transition probabilities and must be recorded in order to construct the time-reversed model. It may be that some of the ν states given by these calculations are identical to those already found. In this case, the sequence including them need not be continued. This process of prepending symbols and calculating transition probabilities is repeated until no unique states are found. Then the ϵ-machine is simply this collection of states $\mathcal{S}^-$, along with the transition matrices that can be derived from the probabilities. It may be that transient states separate from the stationary state exist; depending on the purpose, these may be retained or discarded from the model. Performing the calculations to evaluate each individual state is tedious; the result in the example case is that the recurrent causal states of the reverse ϵ-machine are given by $\nu(10)$, $\nu(100)$, $\nu(1001)$.

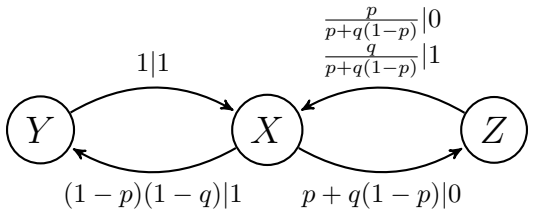


Figure 3: The time-reversed ϵ-machine $\mathcal{M}^-$ of the RnC process

The forward statistical complexity is

$$\begin{aligned} C_\mu^+ &= -\sum_{S\in\mathcal{S}^+} \Pr(S)\lg\Pr(S) \\ &= -\frac{1}{2}\left(\lg\frac{1}{2} + p\lg\frac{p}{2} + (1-p)\lg\frac{1-p}{2}\right) \\ &= -\frac{1}{2}\left(-1 + p\lg p - p\lg 2 + (1-p)\lg(1-p) - (1-p)\lg 2\right) \\ &= 1 - \frac{1}{2}\left(p\lg p + (1-p)\lg(1-p)\right) \end{aligned} \tag{6}$$

Similarly, the reverse statistical complexity is calculated

$$\begin{aligned} C_\mu^- &= -\sum_{S\in\mathcal{S}^-} \Pr(S)\lg\Pr(S) \\ &= -\frac{1}{2}\left(\lg\frac{1}{2} + (1-p)(1-q)\lg\frac{(1-p)(1-q)}{2} + (p+q(1-p))\lg\frac{p+q(1-p)}{2}\right) \\ &= 1 - \frac{1}{2}\left((1-p)(1-q)\lg(1-p)(1-q) + (p+q(1-p))\lg(p+q(1-p))\right) \end{aligned} \tag{7}$$

So

$$\begin{aligned} \Xi &= C_\mu^+ - C_\mu^- \\ &= \frac{1}{2}\left[\left(p\lg p + (1-p)\lg(1-p)\right) - \left((1-p)(1-q)\lg(1-p)(1-q) + (p+q(1-p))\lg(p+q(1-p))\right)\right] \end{aligned} \tag{8}$$

We present plots of both Ξ and h_μ over the space of p, q values for the RnC process in figure 4.

2.2 Definitions of reversibility

We may use the general discrepancy between the statistical complexities of $\mathcal{M}^+$ and $\mathcal{M}^-$ to provide a coarse measure of the degree to which a process is irreversible, which motivates the definition of the *causal irreversibility* Ξ:

$$\Xi = C_\mu^+ - C_\mu^- \tag{9}$$

This raises questions as to the intuition of reversibility. We describe a process $\mathcal{P}$ as *microscopically reversible* if $\mathcal{M}^+(\mathcal{P}) = \mathcal{M}^-(\mathcal{P})$. It is clear that if this is the case, then $\Xi = 0$. However, the

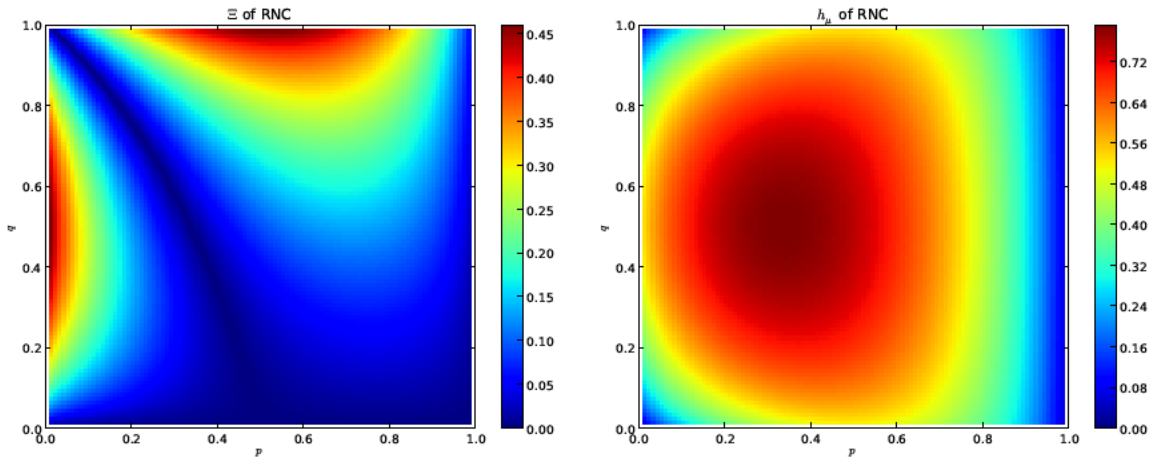


Figure 4: Invariants of the RnC process over the space of p, q values

converse is not true; it is possible that $\Xi = 0$ for a process that is not microscopically reversible. A simple example presented in [3] is that the process generating $\ldots 123123123\ldots$ is not microscopically reversible, as the probability of the word 123 is not the same in the forward ϵ-machine as in the reverse (which generates $\ldots 321321321\ldots$), but $\Xi = 0$, as the entropy of each state in both machines is identical. Furthermore, an automorphism on the alphabet of the reverse machine is sufficient to generate identical processes. Clearly the definition of microscopically reversible is too strict to capture all of the relevant behavior, but causal irreversibility is somewhat imprecise, and obscures a process's structure.

2.3 The $\{N, M\}$-Flower Process

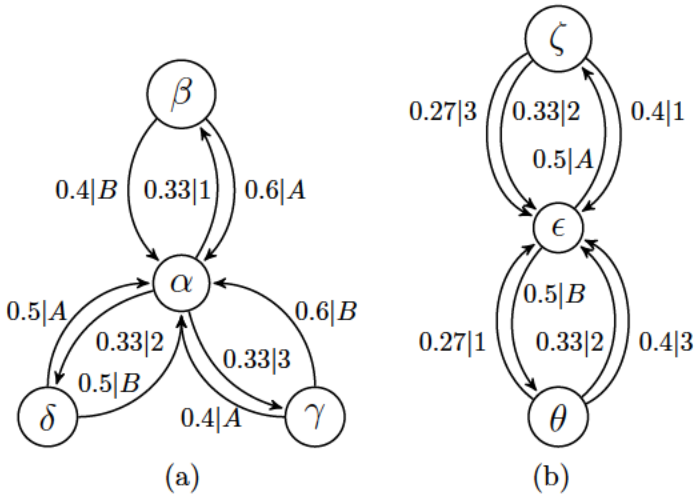


Figure 5: Example forward (a) and reverse (b) ϵ-machines for a $\{4, 3\}$-Flower process. Symbols $\{1, 2, 3\}$ correspond to the three forward "petal" states, while symbols $\{A, B\}$ correspond to the two time-reversed petals.

We seek a class of processes that can more effectively probe these definitions. To that end, we introduce the Flower processes, a family of ϵ-machines taking two parameters, which allows

a fully programmable number of causal states in both the forward and reverse ϵ-machines. The Flower process is structured such that the forward-time process contains a central causal state connected to $N-1$ "petals." There is an outgoing path from the node to each petal, using unique symbols, and the number of return paths to the central state from each petal is equal to the number of time-reversed causal state petals, $M-1$. Thus, an alphabet of size $|\mathcal{X}| = N+M-2$ is necessary for this implementation. If there is degeneracy in the return transition probabilities to the central state in the forward model, these will collapse in the time-reversed model; thus these distributions must be unique to each petal, as can be observed in figure 5. We will use the Flower with uniform "outgoing" distribution (leaving central state) in the forward-time direction. In these circumstances, the number of petals in each state is directly tied to the causal irreversibility.

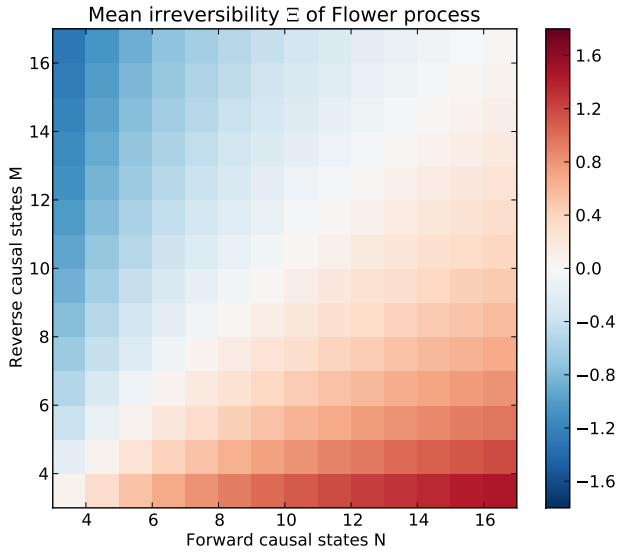


Figure 6: The correlation between irreversibility and N, M is seen to be $\Xi \propto N - M$, or more suggestively, $C_\mu^+ - C_\mu^- \propto |\mathcal{S}^+| - |\mathcal{S}^-|$

Figure 6 shows the mean Ξ of 750 Flowers for every ϵ-machine with $N, M \in \{3, \ldots, 16\}$, with uniform outgoing forward-time distributions and randomly-generated, but unique, incoming probabilities. They show that for the Flower process, the difference in the number of forward and reverse states in large part determines the irreversibility. These results are easy to interpret. Adding states to one direction of the machine boosts the complexity in that direction while having little effect on the complexity in the other. Thus, when the number of states is the same, the complexities are the same. The cost of increasing the number of states in use in the Flower process is the increasing alphabet.

The effect of the "incoming" probability distributions on the irreversibility of the Flower process is the subject of a proof presented in Appendix B. While the process is tedious, the results are illustrated by figure 7. If the transition probabilities are made too extreme (that is, one set at 1 and the others at 0), the machine's topology is altered and it no longer lies within the family of Flower processes. However, as can be seen from the figure and in the proof, the second derivative of Ξ is positive along the space of transition probabilities. Thus, the extremum within this simplex is a minimum, with what we shall refer to as *phase transitions* (that is, alterations to the graph structure of the ϵ-machine) occurring at the boundaries and the minimum.

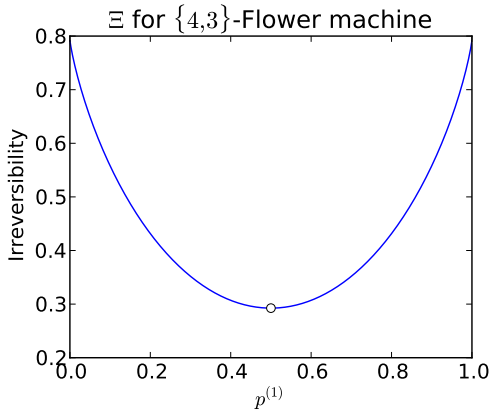


Figure 7: Irreversibility is evidently minimized for a uniform probability distribution, in this case for $p^{(1)} = p^{(2)} = 1.5$, as the number of forward petals is 3 and the number of reverse petals is 2.

3 Topological irreversibility

The existence of phase transitions in ϵ-machine topology at the extrema suggests that the graph structure of a process is more fundamental in determining the irreversibility than the transition probability distributions. We thus turn our attention to structures that can more clearly demonstrate the relevant aspects of a process.

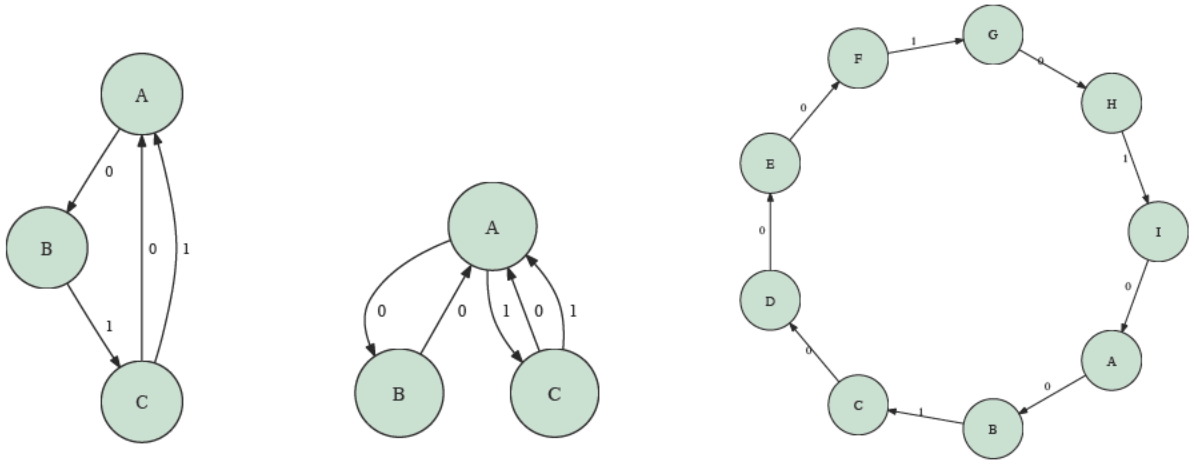
3.1 Finite automata

To do so, it is helpful to venture outside the formalism of the ϵ-machine. We reproduce the definition from automaton theory given in [5]: a *deterministic finite automaton* (DFA) is a 5-tuple $\langle Q, \Sigma, \delta, q_0, F \rangle$, where Q is a finite set of states, Σ is a discrete alphabet, $\delta : Q \times \Sigma \to Q$ is the transition function, q_0 is the start state, and $F \subseteq Q$ is the set of accepting states. To match the notation used in the definition of the DFA with our previous notation: $Q \cong \mathcal{S}$, the causal states, and $\Sigma \cong \mathcal{X}$, the process alphabet. The transition function δ is analogous to the set of transition matrices $\mathcal{T}$, and is interpretable as the act of making an observation; that δ is well-defined corresponds to the property of unifilarity and is the source of the *deterministic* qualifier of the DFA.

We use the concept of a DFA in a restricted sense, discarding some of its utility as a theoretical concept in order to provide a closer abstraction of the ϵ-machine formalism. To that end we set the start state $q_0 = \pi$, representing an initial lack of knowledge of the state of the system. We also set $F = Q = \mathcal{X}$; that is, any causal state is allowed to be the "final" state of the system. These stipulations discard the notion of an initial and final point in time for the system, allowing us to continue treating it as a bi-infinite series of random variables, a process evolving in time by continued application of the δ function on the present state. The advantage over the ϵ-machine formalism is that transition probabilities are discarded; either a particular transition in states is allowed by δ or it is not.

Examples of DFAs are given in figure 8. It should be noted that we now restrict the alphabet to $\mathbb{Z}/2 = \{0, 1\}$. We claim that this induces no loss of generality. Given a process with some alphabet $\mathcal{X}$, a binary encoding can be constructed of the symbols in $\mathcal{X}$, essentially mapping the process to another with an alphabet of $\mathbb{Z}/2$ at the cost of increased complexity in the causal states.

Our definition of causal irreversibility Ξ is not calculable for a DFA, except by assuming some



(a) (b) (c)

Figure 8: Examples of discrete finite automata with binary alphabets, showing similarities to graph presentations of ϵ-machines without transition probabilities

probability distribution not provided by the formalism. We therefore require a new test of irreversibility. We refer to this as *topological reversibility* and define it as the existence of a graph isomorphism between the forward and reverse processes, matching all states and transitions. A lack of such an isomorphism between the forward and time-reversed DFAs is *topological irreversibility*. There is a caveat to this definition, as it is possible for a DFA to have a finite number of states in the forward direction but to have the determinization algorithm of the time reversed machine not converge; that is, there is an infinite number of states in the reversed DFA. We refer to a DFA with this property as *infinitely irreversible*; its reverse cannot be determinized. Each DFA fits into one of these three categories. As an auxiliary component of the research performed for this project we surveyed all topological ϵ-machines (DFAs, as used here) below a certain size in an attempt to gauge the relative incidence of each of these characterizations. The results of that study are presented in Appendix C.

3.2 Semiperiodic DFAs

We restrict our analysis to a particular model class of DFAs. A *periodic DFA* is one which repeats a periodic sequence of some finite length, and has a cyclic graph representation like that shown in figure 8(c). This model class is closed under time reversal, as the automata produced by application of the $\mathcal{T}$ operator are deterministic and the action of $\mathcal{U}$ is trivial. This property is referred to as *counifilarity*. We introduce one feature to this family to produce the class of semiperiodic DFAs, a phase slip transition or *defect* that skips a certain number of states. Because this definition is not found in the literature, we state it precisely: *A semi-periodic DFA (SPDFA) is one built on a periodic DFA, with the addition of n nonoverlapping phase slip, or defect, transitions in the forward direction, of lengths $\{r_1, \ldots, r_n\}$ and spacings $\{l_1, \ldots, l_n\}$ which bypass one or more consecutive states.* Note that our intuition of a directionality ('forward', 'bypass') comes from the periodic backbone word of the DFA. An example of a SPDFA is given in figure 9, and it is reproduced in figure 10, showing the crucial property: though the model is simple, by adding the

defect transitions we can induce topological irreversibility. The specification of the r and l values characterizing a SPDFA assists in identifying topological irreversibility. If the time-reversed r^- and l^- sequences are not cyclic permutations (permuted by the same number of indices) of r^+ and l^+, then the SPDFA is topologically irreversible.

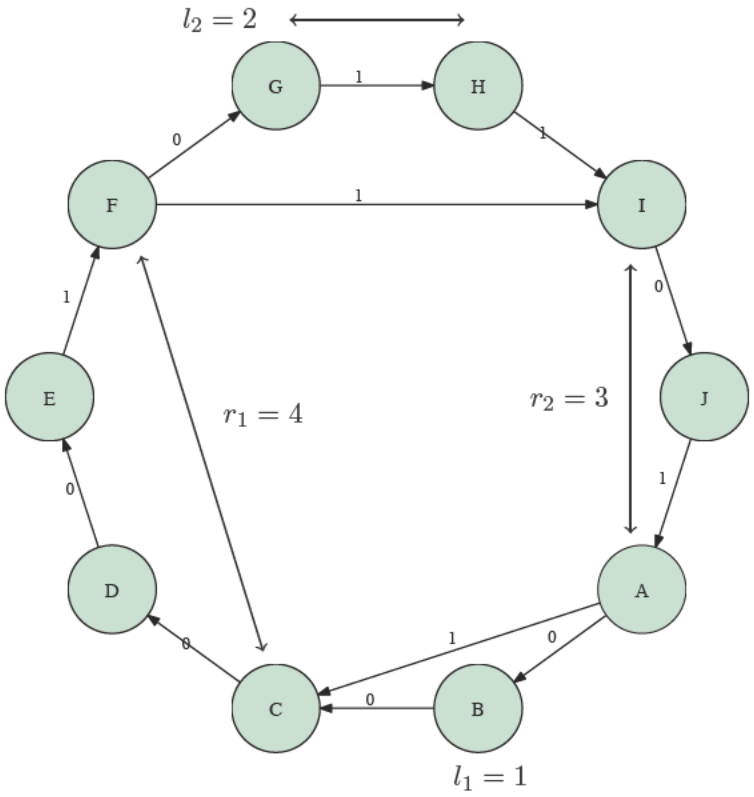


Figure 9: Example SPDFA, showing the specifications of the $l = \{1, 2\}$ and $r = \{4, 3\}$ parameters. Note that $|\mathcal{S}| = \sum_i r_i + \sum_j l_j$.

Under what circumstances is a SPDFA reversible, irreversible, or infinitely irreversible? As defined here, a SPDFA with $n = 1$ is categorically reversible, for reasons that will become clear. The simplest system that shows irreversible behavior is the $n = 2$ case. Therefore, we focus on it. An example of an irreversible $n = 2$ system is shown in figure 10, with the chosen initial phase ϕ_0 indicated. A general property can be seen in this example, namely that the l values, the lengths of the defects, stay constant under time-reversal. This is true as long as the time-reversed DFA remains within the model class. It also shows why a $n = 1$ system must be reversible, as the l value being constant implies that the r value is constant (by the relationship given in figure 9). In the $n > 1$ case, we may attempt to quantify the amount of irreversibility by calculating

$$\Delta r_{\min} = \min r^+ - \min r^- \tag{10}$$

This is a particularly explicit measure in the $n = 2$ case.

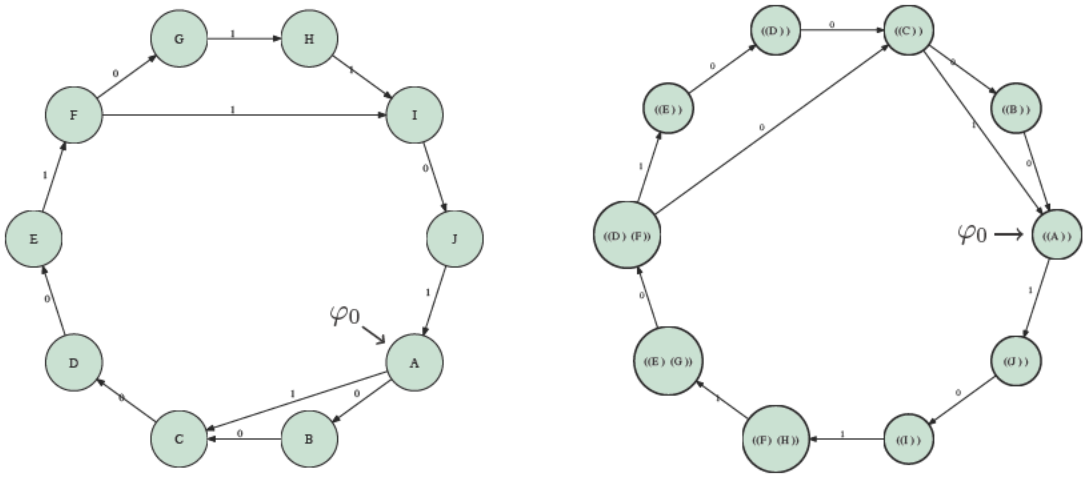


Figure 10: Demonstration of irreversibility of the SPDFA shown in figure 9, with initial phase indicated

3.3 Ensemble descriptions

In order to determine the effect on irreversibility of the r-values and the symbol patterns in the neighborhood of the defects, we focus on a single, long periodic word backbone (60 states) and fixed slip lengths $l = \{1, 1\}$. We then advance the two defects around the word, at each site measuring $\Delta r_{\min}$. For cases where the defects overlap (which by definition breaks the model class) and those where the defects in the time-reversed DFA overlap (again breaking the model class), no value is computed. In this way, by using a single SPDFA we can acquire an ensemble description of irreversible behavior as a function of defect parameters and the neighboring symbols. The results for two words are given in figures 11 and 12, with the axes representing the state of origination of one of the defect transitions, each skipping a single state.

The first graph shows a randomly-generated word which has a plaid structure indicating bands of high and low irreversibility based on the position in the word of each defect. The plot is symmetric about the $x = y$ line, as the defects are identical. If the graph is matched to the word, it can be seen that the strong stripes observed at sites 24 and 29 correspond to the ends two longest blocks of repeated symbols, '00000' (sites 20–24) and '11111' (sites 25–29). However, within the blocks irreversibility remains low. Similarly, sections of very low irreversibility correspond to sites 8–13 and 40–44, subwords consisting of alternating symbols, but no large feature is seen at the end of the section.

The word used for the second graph, figure 12, was created in response to these observations of the previous one, and displays a far higher degree of structure. In particular, the very long section of repeated symbols in the middle of the word results in a similarly very strong site of irreversibility at position 37. The section of alternating symbols that follows shows low irreversibility throughout.

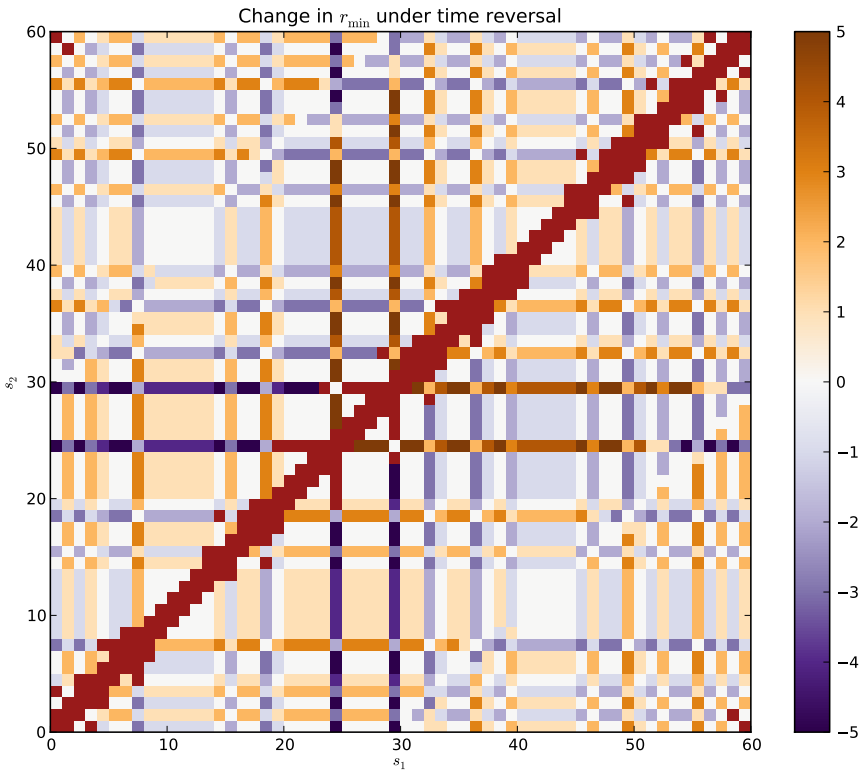


Figure 11: Word 1100100010101011000100000111110001000100101010011101100011100

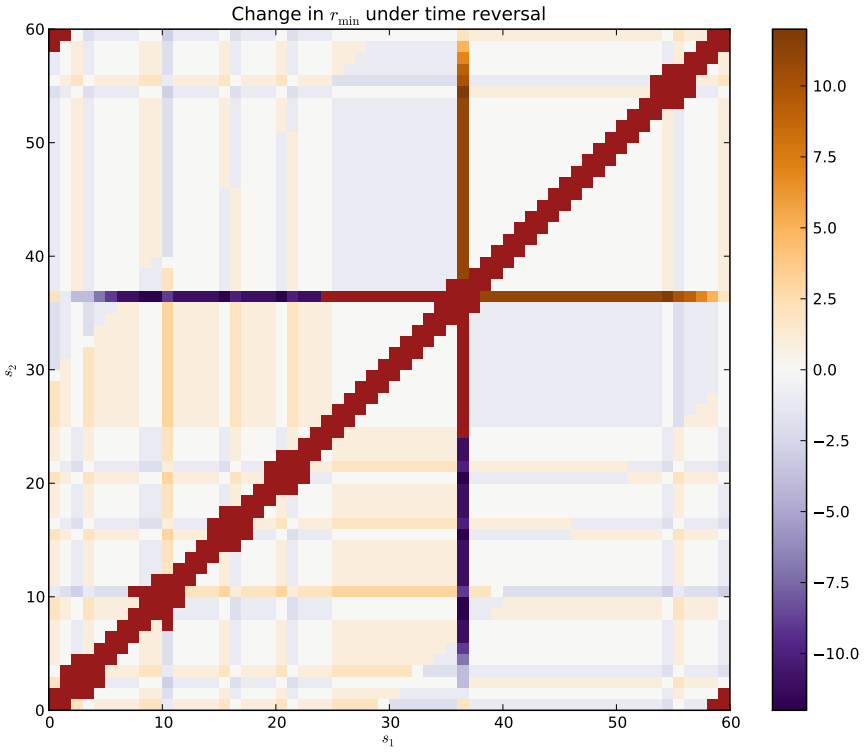


Figure 12: Word 0100101011101010010100101000000000000101010101010101010010100

4 Sources of irreversibility

We propose a local, algorithmic rule that can calculate all of the irreversibility observed in figures 11 and 12. It is as follows: for all defects in the SPDFA which are not locally counifilar,

1. Identify subword that would be encapsulated by time-reversed defect
2. Total shift distance is equal to the number of times this subword was repeated in the periodic word prior to the defect

Why does this rule work? Figure 13 shows the prediction generated by the rule applied to the randomly-generated SPDFA word, and the differential between the actual value and the prediction, showing that it is accurate for all of the SPDFAs which remain in the model class. It can be grasped intuitively by understanding the graphs of semiperiodic structure as a system of pathways. One can traverse the DFA by following any allowed pathway in the forward direction and observing the symbol. Because the process is deterministic, one always knows the current state after observing the symbol; thus, the state of the system and one's knowledge of the state are identical.

Consider now applying the time-reversal operator. The directionality of all pathways flips, and one suddenly cannot make a statement about what state the system is in. One can now traverse the reversed system, but upon encountering a site of nonunifilarity the same symbol is observed for multiple choices of pathway. Thus, it cannot be known which path has been taken, and the state of mind of the observer is a combination of the possible states—in this case, the next state in the periodic word and the one arrived at by skipping over the encapsulated subword. One's knowledge of the state deviates from the state itself. This lack of knowledge (state-mixing) persists until reaching a synchronizing symbol. As long as the encapsulated subword and the word observed by taking the time-reversed defect transition are identical, the state is mixed and there is no distinction made between having taken the defect transition or having advanced in the (reversed) periodic word. Once the first symbol is observed that differentiates these two options, then the path taken becomes known, and we observe the transition to have taken place at this location. This provides the mechanism by which defect transitions are moved through a combination of local counifilarity and repetition of the encapsulated subword, providing the entirety of the irreversibility of the SPDFA class.

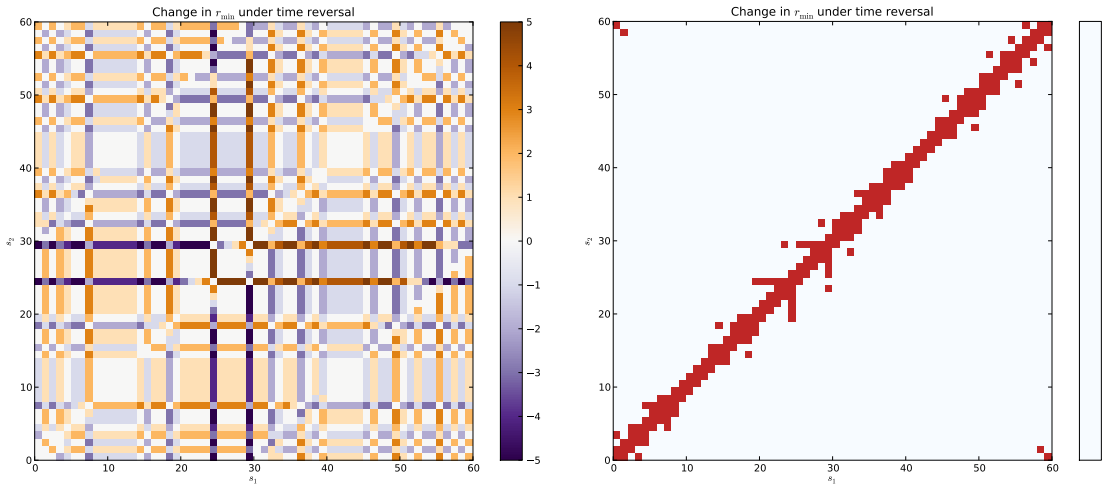


Figure 13: Output and comparison of the SPDFA time-reversal prediction rule given in section 4

5 Conclusion & acknowledgments

We have found using the formalism of the ϵ-machine that irreversibility in a process stems from two distinct sources: the probability distributions and the topology. The system topology is the more fundamental, in that topological irreversibility is a prerequisite for causal irreversibility, an invariant of a process that can be tuned using transition probabilities. The "defect" motif presented for topological ϵ-machines is pattern of irreversibility driven only by the local process structure.

As has been stated, the definition of the semi-periodic DFA provides a very narrow set of circumstances with which to work; the value of understanding irreversibility in this system is not of itself, but as a beginning to understanding other processes. The systems presented here are very close to trivial, but demonstrate the beginnings of behavior that we observe much more generally in larger systems, as shown in Appendix C. This makes it a good beginning, possibly a "zero-order" term, for characterizing time irreversibility as a whole. For example, if a complicated ϵ-machine is structured locally like the semiperiodic model, one could use the rule given here to make a prediction of the locally irreversible behavior. It could even be possible to identify other simple systems and, based on deconstruction into fundamental patterns, to characterize their interaction in a larger ϵ-machine and predict the irreversibility of the entire system. This work may provide the initial step to a project of this nature.

This project was advised by Dr. Jim Crutchfield, whose deep experience with this field was invaluable. John Mahoney expended a great deal of his own time and thought during the summer on helping us out with the specific tasks required as well as planning and managing the overall arc and goals of the project. Chris Strelioff and Ryan James were likewise of great help with learning Python in addition to the excellent CMPy package and the scheduler of the CSC Hive cluster. Finally we wish to acknowledge Rena Zieve and the others associated with the UC Davis REU program in physics, funded under NSF grant PHY-1004848.

List of Figures

References

[1] Crutchfield, J. P. Between order and chaos. *Nature Physics 8* (January 2012), 17–24.

[2] Crutchfield, J. P., and Ellison, C. J. The past and the future in the present, 2010. Submitted (arXiv:1012.0356).

[3] Ellison, C. J., Mahoney, J. R., and Crutchfield, J. P. Prediction, retrodiction, and the amount of information stored in the present. *J. Stat. Phys. 136:6* (2009), 1005–1034.

[4] Ellison, C. J., Mahoney, J. R., James, R. G., Crutchfield, J. P., and Reichardt, J. Information symmetries in irreversible processes. *CHAOS 21:3* (2011), 037107.

[5] Johnson, B. D., Crutchfield, J. P., Ellison, C. J., and McTague, C. S. Enumerating finitary processes. *Theoretical Computer Science 24* (2012), 27–37.

[6] Shalizi, C. R., and Crutchfield, J. P. Computational mechanics: Pattern and prediction, structure and simplicity. *J. Stat. Phys. 104* (2001), 819–881.

A Sample processes and invariants

In order to demonstrate the intuition of the h_μ and C_μ invariants and the state vector π, they are explicitly calculated here for some of the most basic canonical processes. These examples are motivated by those given in reference [3].

Fair coin: In the fair-coin model, each random variable X_i within the time series is IID with the following distribution: $\{(H, \frac{1}{2}), (T, \frac{1}{2})\}$. Because of this property, there is no interaction between any symbols. We therefore expect no complexity to be present because all histories are grouped within a single causal state, α.

The transition matrices $\mathcal{T}$ and stationary causal state distribution π are trivial in this instance, as there is only a single causal state:

$$T = \left[1\right] \qquad T^{(H)} = \left[\tfrac{1}{2}\right] \qquad T^{(T)} = \left[\tfrac{1}{2}\right] \qquad \pi = \left[1\right]$$

Figure 14: Diagram of a fair coin toss

As we are now in possession of $(\mathcal{S}, \mathcal{T})$, we have specified the fair coin's ϵ-machine. Given this information, the predictive complexity of the fair coin is easily calculable:

$$\begin{aligned} C_\mu^+ = H[\mathcal{S}] &= -\sum_{S \in \mathcal{S}} \Pr(S) \lg \Pr(S) \\ &= -(1 \cdot 0) \\ &= 0 \text{ bits.} \end{aligned}$$

It is no surprise that this is the case, as any process with only a single causal state cannot carry any information about its own history. In fact, the history is entirely irrelevant to the fair coin at any time, as it can never deviate from a single state. The entropy rate is

$$\begin{aligned} h_\mu = H[X|\mathcal{S}] &= -\sum_{S \in \mathcal{S}} \Pr(S) \sum_{x, S'} T_{SS'}^{(x)} \lg T_{SS'}^{(x)} \\ &= -1 \left(\frac{1}{2}(-1) + \frac{1}{2}(-1) \right) \\ &= 1 \text{ bit.} \end{aligned}$$

This result, too, is easily intuitively verified. The probability distribution is mixed maximally between the two outcomes, so each new symbol generates a new, fully random, bit of information.

Golden mean: The golden mean process is that which generates all binary strings except those which contain 00 (in this case, BB). It has two states $\mathcal{S} = \{\alpha, \beta\}$, which can be thought of as all histories ending in an A and all of those ending in a B. If the current state S_0 is α, then there is a 50% chance of the next character being an A and 50% chance of it being a B. However, if S_0 is β, the next character will be an A with certainty. This is represented in the transition matrices. Solving for π is straightforward, knowing the eigenvalue to be 1.

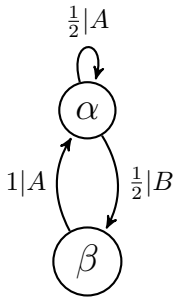


Figure 15: Representation of golden mean process

$$T = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{bmatrix} \quad T^{(A)} = \begin{bmatrix} \frac{1}{2} & 0 \\ 1 & 0 \end{bmatrix} \quad T^{(B)} = \begin{bmatrix} 0 & \frac{1}{2} \\ 0 & 0 \end{bmatrix} \quad \pi = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

The complexity is calculated using equation 2:

$$\begin{aligned} C_\mu^+ = H[\mathcal{S}] &= -\sum_{S \in \mathcal{S}} \Pr(S) \lg \Pr(S) \\ &= -\left[\frac{2}{3} \lg \frac{2}{3} + \frac{1}{3} \lg \frac{1}{3})\right] \\ &= -\left[\frac{2}{3} - \lg 3\right] \\ &= \lg 3 - \frac{2}{3} \approx 0.9183 \text{ bits} \end{aligned}$$

And the entropy rate using equation 3. Note that it is essentially the averaged entropy of all of the causal states. α is maximally distributed between two states and so carries 1 bit of entropy, and occurs $\frac{2}{3}$ of the time in the stationary distribution. β carries no entropy, occurring $\frac{1}{3}$ of the time. Thus, this result is easily predictable.

$$\begin{aligned} h_\mu = H[X|\mathcal{S}] &= -\sum_{S \in \mathcal{S}} \Pr(S) \sum_{x, S'} T_{SS'}^{(x)} \lg T_{SS'}^{(x)} \\ &= -\left[\frac{2}{3}\left(\frac{1}{2}(-1) + \frac{1}{2}(-1)\right) + \frac{1}{3}\left(1(0)\right)\right] \\ &= \frac{2}{3} \text{ bits} \end{aligned}$$

Even process: The even process looks much like the golden mean, but with different labeling. It generates a sequence of B symbols, from which it is impossible to know the causal state, along with some synchronization symbols A, from which the state can be determined. It is the first example considered in which the state of the system is truly "hidden." The decomposition of the transition matrix is slightly different, but the result is identical to that of the golden mean.

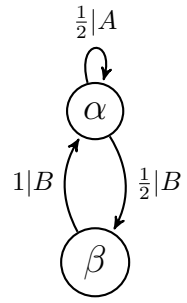


Figure 16: Representation of even process

$$T = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{bmatrix} \quad T^{(A)} = \begin{bmatrix} \frac{1}{2} & 0 \\ 1 & 0 \end{bmatrix} \quad T^{(B)} = \begin{bmatrix} 0 & \frac{1}{2} \\ 0 & 0 \end{bmatrix} \quad \pi = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

Because the state distribution is identical to the golden mean, the statistical complexity is also the same.

$$\begin{aligned} C_\mu^+ = H[\mathcal{S}] &= -\sum_{S \in \mathcal{S}} \Pr(S) \lg \Pr(S) \\ &= -\left[\frac{2}{3} \lg \frac{2}{3} + \frac{1}{3} \lg \frac{1}{3})\right] \\ &= -\left[\frac{2}{3} - \lg 3\right] \\ &= \lg 3 - \frac{2}{3} \text{ bits} \end{aligned}$$

The entropy rate is not guaranteed to be the same as that of the golden mean, as the individual symbol transition matrices are distinct:

$$\begin{aligned} h_\mu = H[X|\mathcal{S}] &= -\sum_{S \in \mathcal{S}} \Pr(S) \sum_{x, S'} T_{SS'}^{(x)} \lg T_{SS'}^{(x)} \\ &= -\left[\frac{2}{3}\left(\frac{1}{2}(-1) + \frac{1}{2}(-1)\right) + \frac{1}{3}\left(1(0)\right)\right] \\ &= \frac{2}{3} \text{ bits} \end{aligned}$$

However, the entropy rate does end up ultimately being the same.

B Derivation and extremization of Ξ for Flower process

B.1 Uniform outgoing forward-time probability distribution

The outgoing forward-time probability distribution affects the irreversibility of the Flower process in the following way. Forward complexity is simply requires application of equation 2. Since all petal states have certainty of transitioning to the central state, its probability in the stationary distribution is $\frac{1}{2}$. Thus

$$\begin{aligned} C_\mu^+ &= -\sum_{i=1}^{N} \Pr(S_i) \lg \Pr(S_i) \\ &= \frac{1}{2} - \sum_{i=2}^{N} \Pr(S_i) \lg \Pr(S_i) \end{aligned}$$

We wish to maximize this function if we are to control the entropy. Intuitively this is given either by creating one petal for which $\Pr(S) \approx 1$, or, by symmetry, by distributing the probability equally amongst all petals. Because $\lim_{x\to 0} x \lg x = \lim_{x\to 1} x \lg x = 0$, if we were to group all of the probability in a single element the Shannon entropy of the petals will vanish. We distribute the probability evenly among petals:

$$\begin{aligned} C_\mu^+ &= \frac{1}{2} - \sum_{i=2}^{N} \Pr(S_i) \lg \Pr(S_i) \\ &= \frac{1}{2} - \sum_{i=2}^{N} \frac{1}{2}\frac{1}{N-1} \lg(\frac{1}{2}\frac{1}{N-1}) \\ &= \frac{1}{2} + \frac{1}{2}(N-1)\frac{1}{N-1} \lg\big(2(N-1)\big) \\ &= \frac{1}{2} + \frac{1}{2}\big[\lg(N-1) + 1\big] \\ &= 1 + \frac{1}{2}\lg(N-1) = 1 + \frac{1}{2}\lg(N_{\text{petals}}) \end{aligned} \tag{11}$$

The forward-time complexity of the evenly distributed Flower process diverges as $\Theta(\lg N)$, whereas the alphabet size diverges as $\Theta(N)$.

For ease of notation, we redefine N and M to be the number of forward and reverse *petal states*, respectively, of a particular ϵ-machine; that is $N = N_{\text{petals}} = N - 1$ and $M = M_{\text{petals}} = M - 1$ (the N and M on the right-hand sides of these equations represent the Flower parameters and the left-hand sides their usage here). Each of the forward-time petal states S_i has an incoming transition probability $p_i^{(x)}$ associated with each reverse symbol x. Time-reversing the system, after transposing we have to normalize each of these probabilities by the corresponding outgoing probability, which we know in this case to be $\frac{1}{N}$ since the "incoming" forward-time probability distribution is uniform. Analysis of the mixed-state presentation is aided by the form of the state transition matrix for the Flower process. It is, in the general case, ($p^{(j)} = \sum_i^N p_i^{(j)}$)

$$\widetilde{T} = \begin{bmatrix} 0 & \frac{p^{(1)}}{N} & \cdots & \frac{p^{(M)}}{N} \\ \mathbf{1}_N & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}$$

Where boldface numerals indicate column vectors. The decomposition is simple: the i^{th} "outgoing" state corresponds to a predominantly empty matrix holding the i^{th} entry in the $\mathbf{1}$ column vector.

The "incoming" states are the same, but with each containing its total probability at its position j along row 0. Using these matrices to form the mixed state yields the following stationary distribution without difficulty, since again we know that after transitioning to a "petal" state, a return to the central (reverse) causal state is certain:

$$\pi^- = \begin{bmatrix} \frac{1}{2} & \frac{p^{(1)}}{2N} & \cdots & \frac{p^{(M)}}{2N} \end{bmatrix}$$

Therefore the reverse statistical complexity is

$$\begin{aligned} C_\mu^- &= -\sum_{i=1}^{N} \Pr(S_i) \lg \Pr(S_i) \\ &= \frac{1}{2} - \sum_{j=1}^{M} \frac{p^{(j)}}{2N} \lg \left(\frac{p^{(j)}}{2N} \right) \end{aligned} \tag{12}$$

And finally, adding equations 11 and 12, the causal irreversibility of a uniformly forward-distributed $\{N, M\}$-Flower machine is

$$\Xi = \frac{1}{2} + \frac{1}{2} \lg N + \sum_{j=1}^{M} \frac{p^{(j)}}{2N} \lg \left(\frac{p^{(j)}}{2N} \right) \tag{13}$$

In order to maximize the uniform-outgoing Flower process, we differentiate equation 13 with respect to a vector $\mathbf{p}$ composed of the incoming transition probabilities $\mathbf{p} = \begin{bmatrix} p^{(1)} & p^{(2)} & \cdots & p^{(M)} \end{bmatrix}$, where $p^{(j)} \in (0, 1) \; \forall \; j \in \{1, \ldots, M\}$. This vector, being stochastic, must satisfy the law of total probability, where the vector sums to N instead of to 1 because the values given in $\mathbf{p}$ are distributed over N forward causal petals, each of which must sum to 1:

$$|\mathbf{p}| = \sum_{j=1}^{M} p^{(j)} = N \tag{14}$$

This gives the constraint on probabilities as a relationship between all variables $p^{(j)}$, namely that they sum to N. From this we can deduce a rule for partial derivatives:

$$\frac{\partial p^{(j)}}{\partial p^{(i)}} = \begin{cases} 1, & i = j \\ -1, & i \neq j \end{cases} \tag{15}$$

In order to find extrema of Ξ subject to the constraint defined in equation 14 we use the method of Lagrange multipliers. The Lagrangian is

$$\begin{aligned} \Lambda &= \nabla \Xi + \lambda \nabla \left(N - |\mathbf{p}| \right) \\ &= \nabla \sum_{j=1}^{M} \frac{p^{(j)}}{2N} \lg \left(\frac{p^{(j)}}{2N} \right) - \lambda \nabla \sum_{j=1}^{M} p^{(j)} \end{aligned} \tag{16}$$

Taking the Lagrangian to be stationary gives, $\forall\, k \leq M$,

$$\lambda \tfrac{\partial}{\partial p^{(k)}} \sum_{j=1}^{M} p^{(j)} = \tfrac{\partial}{\partial p^{(k)}} \left(\sum_{j=1}^{M} \frac{p^{(j)}}{2N} \lg \frac{p^{(j)}}{2N} \right)$$

$$\lambda \left(\sum_{j\neq k}^{M} \tfrac{\partial p^{(j)}}{\partial p^{(k)}} + \tfrac{\partial p^{(k)}}{\partial p^{(k)}} \right) = \frac{1}{2N} \tfrac{\partial}{\partial p^{(k)}} \left(\sum_{j\neq k}^{M} p^{(j)} \lg \frac{p^{(j)}}{2N} + p^{(k)} \lg \frac{p^{(k)}}{2N} \right)$$

$$\lambda \left[(M-1)(-1) + 1 \right] = \frac{1}{2N} \left[\sum_{j\neq k}^{M} \left(\tfrac{\partial p^{(j)}}{\partial p^{(k)}} \lg \frac{p^{(j)}}{2N} + p^{(j)} \tfrac{\partial}{\partial p^{(k)}} \lg \frac{p^{(j)}}{2N} \right) + \left(\tfrac{\partial p^{(k)}}{\partial p^{(k)}} \lg \frac{p^{(k)}}{2N} + p^{(k)} \tfrac{\partial}{\partial p^{(k)}} \lg \frac{p^{(k)}}{2N} \right) \right]$$

$$(2-M)\lambda = \frac{1}{2N} \left[\sum_{j\neq k}^{M} \left(p^{(j)} \frac{1}{\ln 2} \tfrac{\partial \ln p^{(j)}}{\partial p^{(k)}} - \lg \frac{p^{(j)}}{2N} \right) + \left(\lg \frac{p^{(k)}}{2N} + p^{(k)} \frac{1}{\ln 2} \tfrac{\partial \ln p^{(k)}}{\partial p^{(k)}} \right) \right]$$

$$2N(2-M)\lambda = \sum_{j\neq k}^{M} \left(p^{(j)} \frac{1}{\ln 2} \tfrac{\partial \ln p^{(j)}}{\partial p^{(j)}} \tfrac{\partial p^{(j)}}{\partial p^{(k)}} - \lg \frac{p^{(j)}}{2N} \right) + \left(\lg \frac{p^{(k)}}{2N} + \frac{1}{\ln 2} \right)$$

$$2N(2-M)\lambda = \sum_{j\neq k}^{M} \left(-\lg \frac{p^{(j)}}{2N} \right) + \lg \frac{p^{(k)}}{2N} + \frac{2-M}{\ln 2}$$

$$(2-M) \left(2N\lambda - \frac{1}{\ln 2} \right) = \lg \frac{p^{(k)}}{2N} - \lg \prod_{j\neq k}^{M} \frac{p^{(j)}}{2N}$$

$$\lg \frac{p^{(k)}}{\prod_{j\neq k}^{M} p^{(j)}} = (2-M) \left(2N\lambda - \frac{1}{\ln 2} \right)$$

$$\frac{p^{(k)}}{\prod_{j\neq k}^{M} p^{(j)}} = C \tag{17}$$

That the quantity on the left-hand side of equation 17 is equal to the constant C for all values of k implies that $p^{(m)} = p^{(n)} = \frac{N}{M}\ \forall\, m, n < M$ is the extremum of the function. However, we note that $\nabla^2 \Xi$ is positive for all values of $\mathbf{p}$. Thus, the maximum of the function must occur on the boundary; symmetry again dictates that the optimal case is to set $p^{(k)} = 1$ for some k and have $p^{(j)} = 0$ for all others, but we cannot achieve this, as the topology of the Flower machine breaks when we set any $p = 0$ and we enter a new regime. Therefore, this value of Ξ is a supremum. We can approach this by setting $p^{(k)} \approx 1$; $\forall\, j \neq k$, $p^{(j)} \approx 0$. The behavior of Ξ in response to changes in probability is demonstrated in the simple case of a Flower machine having two time-reversed petals in figure 7 in section 2.3.

B.2 General case

We wish to eliminate the assumption that a uniform outgoing distribution maximizes causal irreversibility. The solution will not be given here, but the form of Ξ is derived. We impose the convention that i be used to index forward-time states summed over N, as in the probabilities $p^+_{(i)}$, $i \in \{1, \ldots, N\}$, where N is the number of forward petal states, and j be used to index reverse-time states summed over M, as in the forward-time "incoming" probabilities $p^+_{(ji)}$, $i \in \{1, \ldots, N\}$, $j \in \{1, \ldots, M\}$, where M is the number of reverse petal states. The time-reversed probabilities are represented as $p^-_{(j)}$ and $p^-_{(ij)}$, $i \in \{1, \ldots, N\}$, $j \in \{1, \ldots, M\}$. If we relax the requirement that the Flower process possess uniform outgoing distribution probabilities, the irreversibility takes the following general form given in terms of forward-time probabilities:

$$\begin{aligned}
\Xi &= C^+_\mu - C^-_\mu \\
&= \left(-\sum_i^{N+1} \Pr(S^+_i) \lg \Pr(S^+_i) \right) - \left(-\sum_j^{M+1} \Pr(S^-_j) \lg \Pr(S^-_j) \right) \\
&= \frac{1}{2} \left(1 - \sum_i p^+_{(i)} \lg(p^+_{(i)}) \right) - \frac{1}{2} \left(1 - \sum_j \left(\sum_i \frac{p^+_{(ji)}}{p^+_{(i)}} \right) \lg \left(\sum_i \frac{p^+_{(ji)}}{p^+_{(i)}} \right) \right) \\
&= \frac{1}{2} \left(\sum_j \left[\left(\sum_i \frac{p^+_{(ji)}}{p^+_{(i)}} \right) \lg \left(\sum_i \frac{p^+_{(ji)}}{p^+_{(i)}} \right) \right] - \sum_i p^+_{(i)} \lg(p^+_{(i)}) \right)
\end{aligned} \tag{18}$$

We wish to maximize this quantity via the sets of parameters $\{p^+_{(i)}, p^+_{(ji)}\}$ subject to constraints originating from the law of total probability.

$$\begin{aligned}
\sum_i p^+_{(i)} &= 1 \\
\sum_j p^+_{(ji)} &= 1
\end{aligned}$$

We note the following differential relations, easily derived from the previous equations.

$$\frac{\partial p^+_{(k)}}{\partial p^+_{(i)}} = \begin{cases} 1, & i = k \\ -1, & i \neq k \end{cases} \qquad \frac{\partial p^+_{(mk)}}{\partial p^+_{(ji)}} = \begin{cases} 0, & m \neq j \\ 1, & m = j,\ k = i \\ -1, & m = j,\ k \neq i \end{cases}$$

The Lagrangian for the general case is

$$\Lambda = \nabla \Xi + \lambda_1 \nabla \left(1 - \sum_i p^+_{(i)} \right) + \lambda_2 \nabla \left(1 - \sum_j p^+_{(ji)} \right)$$

Because the constraining functions are functions only of one of the sets of probabilities, they will be treated independently in the gradient. The stationary points of Λ for all probabilities $p^+_{(mk)}$, $m \leq M$, $k \leq N$ satisfy

$$\lambda_2 \sum_j \frac{\partial p^+_{(ji)}}{\partial p^+_{(mk)}} = \frac{1}{2} \sum_j \left[\frac{\partial}{\partial p^+_{(mk)}} \left(\sum_i \frac{p^+_{(ji)}}{p^+_{(i)}} \right) \lg \left(\sum_i \frac{p^+_{(ji)}}{p^+_{(i)}} \right) + \left(\sum_i \frac{p^+_{(ji)}}{p^+_{(i)}} \right) \frac{\partial}{\partial p^+_{(mk)}} \left(\lg \left(\sum_i \frac{p^+_{(ji)}}{p^+_{(i)}} \right) \right) \right]$$

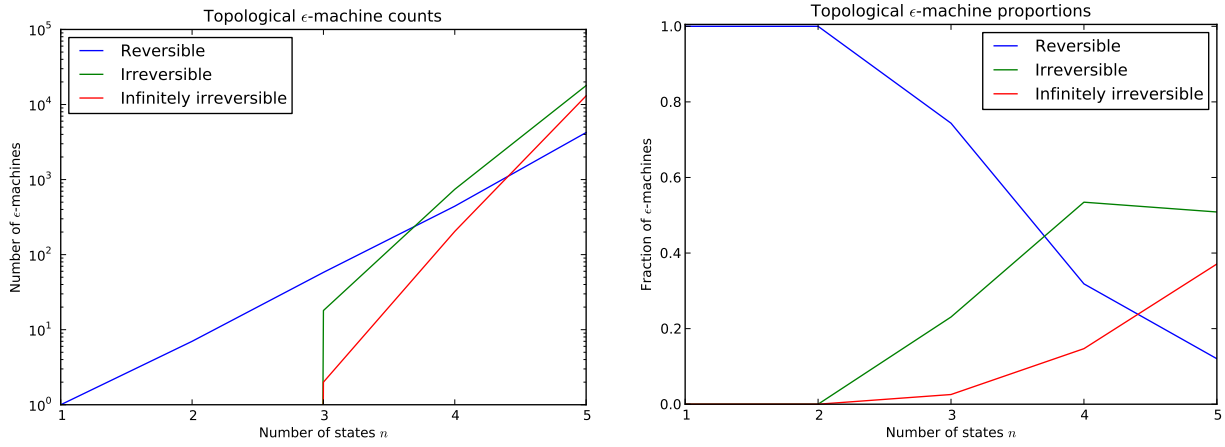
$$\begin{aligned} 2(2-M)\lambda_2 &= \sum_j \left[\sum_i \left(\frac{1}{p^+_{(i)}} \frac{\partial p^+_{(ji)}}{\partial p^+_{(mk)}} \right) \lg \left(\sum_i \frac{p^+_{(ji)}}{p^+_{(i)}} \right) + \frac{1}{\ln 2} \sum_i \left(\frac{\partial p^+_{(ji)}}{\partial p^+_{(mk)}} \frac{1}{p^+_{(i)}} \right) \right] \\ &= \sum_i \left(\frac{1}{p^+_{(i)}} \frac{\partial p^+_{(mi)}}{\partial p^+_{(mk)}} \right) \lg \left(\sum_i \frac{p^+_{(mi)}}{p^+_{(i)}} \right) + \frac{1}{\ln 2} \sum_i \left(\frac{\partial p^+_{(mi)}}{\partial p^+_{(mk)}} \frac{1}{p^+_{(i)}} \right) \\ &= \left[\lg \left(\sum_i \frac{p^+_{(mi)}}{p^+_{(i)}} \right) + \frac{1}{\ln 2} \right] \left[\frac{1}{p^+_{(k)}} - \sum_{i \neq k} \frac{1}{p^+_{(i)}} \right] \end{aligned}$$

The resulting equation is of the form $C = f_m g_k$. Since this holds (with the same constant) for all m and k, which may be varied independently of one another, then both functions must be constant in m and k. This leads to the conclusions that $p^+_{(i)} = \frac{1}{N} \, \forall \, i \leq N$, and $\sum_i \, p^+_{(ji)}$ is identical $\forall \, j \leq M$. These are precisely the conditions applied previously to minimize the complexity of the Flower process.

Again by symmetry, finding the maximum of Ξ requires finding a point that lies on the boundary of the product space of the simplices, which the method of Lagrange multipliers cannot do. It may require numerical solution.

C Survey of topological ϵ-machines

The class of topological ϵ-machines is composed of ϵ-machine structures, stripped of transition probabilities. A method for enumerating all such machines was given in [5] and implemented in the CMPy package for Python, allowing us to perform an exhaustive survey up to a given number of causal states n. The number of processes increases exponentially with the number of states, making this survey a difficult task even at relatively low n. The survey was performed in parallel on the UC Davis CSC Hive cluster. Each ϵ-machine was time-reversed and categorized as reversible, finitely irreversible, or infinitely irreversible based on whether the mixed-state process converges. The results, given in figure 17 show that, while processes with low numbers of states are mostly reversible, with increasing states the proportion of irreversible and infinitely irreversible machines grows. It is impossible to make definite statements given the result, but it could be the case that the proportion of ϵ-machines which are reversible dwindles with larger numbers of states, while the proportion of finitely irreversible ϵ-machines is high in the mesoscale but either drops at high n or equalizes with the proportion of infinitely irreversible ϵ-machines, which suggests and upward trend at the high-n edge of figure 17(b).



(a) (b)

Figure 17: Results from topological ϵ-machine survey, with number of states given on a semilog plot (a) and the relative frequency of each category (b)